Taking Stock of the Genome Project

With Francis Collins at the helm, the Human Genome Project at NIH is lobbying for more money to finish its ambitious job

Francis Collins has never been one to avoid tough jobs. Collins, who was lured to the National Institutes of Health (NIH) last spring to head the National Center for Human Genome Research, has hunted down a string of elusive genes. He tackled cystic fibrosis, joining the team that eventually pulled out the faulty gene and has since been working on a therapy. He co-identified the gene that causes neurofibromatosis 1, and was part of the collaborative group that finally tracked down the gene involved in Huntington's disease earlier this year. Collins and others are now pursuing a breast cancer susceptibility gene—a discovery that could have vast medical, social, and ethical implications.

But in his new job, Collins is confronting a bigger challenge than he's ever faced before. He is taking over the reins of the genome project, formerly headed by Nobel laureate and double helix codiscoverer James Watson, at a critical time. The international program, which is jointly funded here by the Department of Energy (DOE) and NIH, needs a considerable infusion of new dollars if it is to meet its ambitious goal of determining the sequence of all 3 billion base pairs of the human genome by 2005. Yet the project is already getting about $165 million a year, and money for science is scarce. "When you see all the programs that have lots of support taking cuts, it is very difficult to argue successfully for a ramp up of an innovative new program," concedes Collins. "It is a real problem."

Yet funding is only one of myriad issues Collins faces. He and numerous advisers in the genome community have just finished an intensive review of the first few years of the program, taking stock both of its successes and its failures. Based on those findings, they have crafted a new 5-year plan to guide the project through 1998 (see page 43). In some areas, there seems to be a remarkable scientific consensus on where the project ought to go. In particular, everyone agrees there should be a concerted push on DNA sequencing because a lack of new technologies points to a major bottleneck in the years ahead. Essential, too, is increased emphasis on new software and hardware for collecting, disseminating, and analyzing the sequence data (see p. 47). Also in vogue is a shift away from the chromosome as the unit of analysis to a new focus on smaller regions of biological interest—a change that will influence which genome centers survive and which new ones get funded.

On other issues, there's less consensus, however. Researchers are divided over how much effort the project should put into tracking down genes, for example. A program devoted to the ethical, legal, and social implications of the project, known as ELSI, has come under criticism from disgruntled scientists, who want a greater push on public education, and from others who think ELSI should be stepping boldly into policy making. Then there's the ongoing dispute over the propriety of patenting gene fragments.

PAULINE LUBENS/DETROIT FREE PRESS

Facing the challenge. New project head Francis Collins.

So far, however, Collins seems to have the support of the genome community, which he will certainly need. Collins, too, apparently has the support of his bosses in the Department of Health and Human Services (HHS), where a proposal is now wending its way up to create another institute for NIH—a National Institute of Genomics and Medical Genetics, with Collins as its director.

First the good news

By one key measure, Collins has inherited a project in good shape: In terms of meeting its first priority—developing genetic linkage maps of the human genome and those of several model organisms—the project is coming in ahead of schedule and under cost. "We have excellent genetic maps of mouse and human even faster than expected," says Maynard Olson of the University of Washington. Olson attributes much of the success to Jean Weissenbach and colleagues at the Centre d'Etude de Polymorphisme Humain (CEPH) in Paris. Genetic linkage maps, which consist of a series of signposts—usually highly variable pieces of DNA—arrayed along the chromosome, are particularly useful for finding the rough location of disease genes.

A second type of genome map, the physical map, if not ahead of schedule is at least on target. Physical maps are actual assemblages of DNA clones, lined up in the same order as they appear on the chromosome. The ultimate goal is to go from these maps to pulling out genes and sequencing them. At this stage, however, the resolution of these maps is less than ideal. The markers are spaced on average every 300 kilobases, as opposed to the original goal of a 100-kilobase resolution map by 1995.

Even so, the increasingly sophisticated maps and resources, such as DOE's chromosome-specific collections of clones, have speeded the isolation of genes involved in numerous diseases, including Fragile X, Huntington's, and colon cancer. Studies of these genes have, in turn, revealed fascinating genetic mechanisms, such as the trinucleotide repeat mutations that lie at the heart of Fragile X, myotonic dystrophy, Huntington's, and who knows how many other diseases.

Mortgaging the future

But progress on the maps has come at a cost. While mapping is ahead of the schedule originally set in 1991 in the first 5-year plan, sequencing lags behind. Although sequencing speed has risen over the past few years and the cost per base pair has dropped, Collins and others say a 100-fold improvement in speed is still needed if the project is to meet its goal of knocking off the entire human genome by 2005. Indeed, it was partly concerns about the sluggish progress on sequencing that prompted Collins, David Galas, who, until he recently left for Darwin Molecular Technologies in Seattle, oversaw the genome project for DOE, and numerous advisers to revisit the 5-year goals over the spring and summer. The roadblock in sequencing is now money, says Collins. "Good ideas are going begging," he says.

The problem is that the budget has not increased as fast as the project's creators recommended. When biology's first megaproject was planned, a committee of the National Research Council in 1988 concluded that it would take 15 years and cost about $3 billion, or $200 million a year, to pull it off. Although those numbers have withstood repeated scrutiny, asserts Collins, the $200

million has failed to materialize. The combined NIH and DOE budget remained at roughly $165 million from 1992 to 1993, when, adjusted for inflation, it should be at $219 million, says Collins. "We are now being asked to do the project at 75% of the funding that the NRC said it should cost." The upshot, he says, is that most of the money went into mapping, and the "revolutionary sequencing techniques envisioned earlier simply have not materialized. We have mortgaged part of our future."

The new plan calls for $100 million a year exclusively for sequencing technologies, thereby bringing the total budget up to the $200 million equivalent originally recommended. What happens if they don't get it? "The simple answer," says Collins, "is that we are probably not going to be able to make that timetable." And he predicts the consequences would be grim, both in terms of delayed medical benefits and a loss of U.S. biotechnology competitiveness.

What's more, even with the full funding, meeting the sequencing goal will still be a "stretch," concedes Collins. He and others predict that the job will probably have to be done with incremental improvements in today's sequencing technology, based on gel electrophoresis, rather than with glitzy new approaches such as mass spectrometry or atomic force microscopy. What's needed now, Collins and others agree, is automation that will dramatically lower the amount of labor needed to sequence and thereby cut the costs—something that "the genome project has failed to get a grip on," asserts Olson. Until now, people have largely focused on automating individual steps of the sequencing process, says Collins—for instance, building better sequencing machines, or robots for DNA preparation, or software that can analyze and assemble clones in order. Now the focus is shifting toward automating all the steps as a unit so that no one step is rate limiting.

But that poses a tough question about the sequencing budget. We could put all of our eggs into automating current sequencing methods, which we know will work, says Collins. But what then, he asks, about the "blue sky revolutionary ideas" that don't get funded because of the budget crunch—and that could make all the difference?

Despite the slow progress, there is little sentiment at this stage for abandoning the goal of all-out sequencing, Collins says. The biological insights emerging from the few large-scale sequencing efforts, such as those on the nematode *Caenaorhabditis elegans*, are just too alluring. Comparative analyses have revealed, for example, a remarkable similarity in genes shared across species.

But some thought is being given to a shortcut called one-pass sequencing. The original plan calls for sequencing the whole genome several times, to ensure an error rate of 0.001%. "Suppose we try one-pass coverage with 1% error rate but it only costs one-tenth as much?" asks Collins. The idea, then, would be to return to the really interesting regions and sequence them again.

Recruiting new bodies

Aside from cold cash, the sequencing effort will need more warm bodies if it is going to be done on time. Additional groups will have to get involved in large-scale projects, says Collins, tackling a sizable chunk of a chromosome, say, a megabase a year. Right now, Collins can count on one hand the groups that have that capability. One way to recruit sequencers and convince them to scale up is to give them some interesting biology to work on. And that meshes nicely with another change in the new 5-year plan.

Most of the existing genome centers were built around analyzing particular chromosomes. But now there's a shift away from chromosomes to focus instead on regions of biological interest—particularly regions several megabases long that seem to have functional, structural, and evolutionary significance. "It's biology that should dictate the size" of the unit sequenced, says Olson. And to him, that is good news, because it means the genome project will be "more biologically driven, with more room for creativity, and that the centers won't have to be so huge."

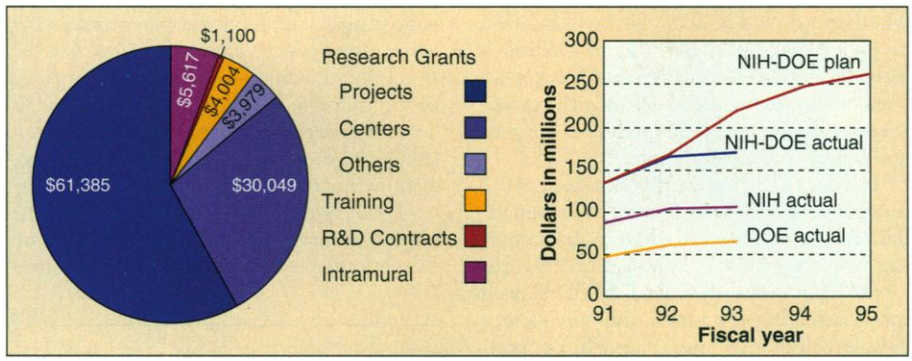
Putting genes on the map

Another major shift in emphasis under Collins, perhaps less universally embraced, is a goal of placing the 100,000 human genes on the maps. Gene identification was always an implicit goal of the project, insists Collins, though it was never stated explicitly, perhaps because of its difficulty. Now several new techniques make gene finding easier, he says, and the annotated map will be far more useful in helping investigators identify disease genes.

Articulation of this goal is also sending an explicit message to the private sector, where enormous efforts—"probably more than we know about," asserts Collins—are under way to partially sequence all the expressed genes or cDNAs. Pending a decision of the patentability of these gene fragments, most of these sequences are being kept secret, says Collins, and the genes are not being put on the map.

But this new emphasis on gene identification is raising questions about what the goal of the genome project really is. As first envisioned, it was to build the infrastructure—the maps and tools—to prepare for the biology of the 21st century, leaving the gene discovery for others. Will Collins, the avid gene hunter, shift the focus too much in the direction of looking for disease genes, especially when technology development is suffering? That possibility worries Leroy Hood, head of the molecular biotechnology department at the University of Washington and also a large-scale sequencer. Hood sees no fundamental role for disease gene hunting in the project, because "there is simply not enough money to go around."

Collins actually agrees, asserting that "building the infrastructure is still our first priority," and pointing out that he means annotating the map with all genes, not just disease genes. But, he adds, "the reason the public pays and is excited—well, disease genes are at the top of the list. We can't take on the entire field of finding genes, but I will be pleased if the project catalyzes it along the way." Collins suspects that some of these concerns may reflect apprehensions about the intramural program he is launching at NIH, which will have a decided focus on disease genes and indeed gene therapy. But, he insists, "the creation of an intramural program with a strong applied focus does not change the extramural program."



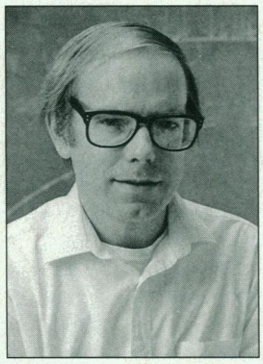
Where it goes. The chart *(left)* shows the breakdown of the NIH genome project budget for FY 1993 (in thousands of dollars). On the right is the funding since 1991.

ELSI at a crossroads

From the start, Watson and others realized that the information garnered from the genome project could be misused—in denying health insurance, for instance. For that reason he set up an ethics program and promised that it would receive at least 3% of genome project funds. It now receives about 5%. But now, having spent 4 years and $20 million,

the ELSI program is at a critical juncture, with numerous critics wondering what it has produced. Asked Olson at a meeting this summer: "Why don't we have any visible progress toward a federal genetics privacy law 3 years into the program?"

Most of what ELSI has done has been to define the high-priority issues that require urgent attention, both through research grants, such as pilot studies on screening for cystic fibrosis, and, most visibly, a series of academic meetings. It is these meetings—where often the same cast of characters debate the same issues—that have taken most of the heat. Some bench scientists are openly fed up. "We've had enough of this Hastings Center stuff," says one. This summer several advisers to NIH and DOE's genome projects complained that ELSI was too divorced from the science and that it was time for them to quit talking and start doing something, though opinion is divided on whether it should be active public education or policy making or both. Even Collins, a staunch supporter, concedes that "people are tired of another venue of defining the issues. It is time to move on and produce some general policy recommendations."



Maynard Olson. Where is a genetics privacy law?

But does ELSI have either the clout or the independence to develop policy recommendations that anyone will listen to? In a report last year, the House Committee on Government Operations described ELSI as well intentioned but too low in the bureaucracy to be effective at setting policy. The committee recommended that an independent body be created to review the ethical, legal, and social implications of the genome project. An upcoming report from the Office of Technology Assessment is also expected to support the notion of an independent bioethics commission.

While Collins maintains that ELSI is independent enough to serve as watchdog, he also says he has no problems with a new commission, provided its budget doesn't come out of ELSI funds. ELSI would still have an indispensable role to play in catalyzing research on ethical issues—particularly on the proper introduction of genetic screening for such common diseases as breast or colon cancer, he says. Other critical roles are educating the public about genetics and its implications and developing policy options for any new commission to consider.

One of the most pressing issues Collins will confront is what type of genetic information can be patented. Hood characterizes this as "the most confused and potentially challenging issue we face. Whether we like it or not, Craig Venter opened up a whole series of questions that are legitimate." Hood is referring to former NIH scientist Craig Venter, now at The Institute for Genomic Research outside Washington, D.C., who sequenced thousands of gene fragments of unknown function. NIH kicked off a furor by applying for patents on the fragments. When the Patent Office rejected the application last year, HHS appealed that decision. What happens next will depend, in large part, on how aggressively incoming NIH Director Harold Varmus pushes the appeal. Collins is hoping the NIH patent will soon be disallowed. Until then, he says, massive cDNA sequencing is proceeding apace in the private sector and those sequences are not being made public. "It is bad for the project."

But even a resolution of this one patent application will not address the other questions Venter's approach has raised. While there seems to be near-universal agreement that gene snippets should not be patentable, what about entire genes, or entire regions, like the T cell locus, which encodes genes involved in the immune response, ponders Hood. "If we complete the sequence of the T cell locus, should we be allowed to patent all the genes even without knowing their function?"

Tricky questions, too, are arising from the myriad of new startup companies based on the genome project, which Collins sees as a sign that the genome project is indeed succeeding (*Science*, 15 January, p. 300). But such success is "twoedged," he concedes, as many of the principals are heavily involved in NIH-funded genome centers. To name a few: David Cox and Rick Myers of Stanford, along with Dennis Drayna, formerly of Genentech, just founded Mercator Genetics Inc., in Menlo Park, and Eric Lander of the Whitehead Institute and Daniel Cohen of CEPH are "founding scientific advisers" at Millennium in Cambridge. The obvious concern is that some investigators might use the resources developed with public funds for their own proprietary interest. Yet another question is how will the genome project receive impartial advice when nearly everyone has a financial stake in it? "We must all be willing to sit under a hot light," says Cox. And Collins, who rid himself of all commercial ties to take this job, vows to keep a tight watch.

Once Collins settles into his new job, he says that he plans to spend about one-third of his time doing science in the intramural lab. For now, he has his hands full moving his lab from Ann Arbor and recruiting some of the nation's top geneticists to join him at NIH. As he dashes from city to city, lobbying for a budget increase on Capitol Hill one day, attending a thesis defense in Ann Arbor on another, Collins seems energized, not cowed, by the challenges that face him. "I like intensity," says Collins, who insists he has yet to regret, even for a moment, his decision to take this job.

–Leslie Roberts

AWARDS

National Science, Technology Medalists

This week the White House awarded the National Medal of Science and the National Medal of Technology to 17 scientists, mathematicians, and engineers. The recipients, three of whom are Nobelists, were honored by President Bill Clinton at a ceremony in the Rose Garden.

National Medal of Science

Biological sciences: Donald J. Cram, University of California, Los Angeles, for research on the chemical foundations of molecular recognition; Daniel Nathans, Johns Hopkins University, for seminal work in molecular genetics; and Salome G. Waelsch, Albert Einstein College of Medicine, for a lifetime of work on developmental genetics;

Chemistry: Norman Hackerman, president emeritus, Rice University, for contributions in electrochemistry and education;

Engineering: Alfred Y. Cho, AT&T Bell Labs, for development of molecular beam epitaxy;

Mathematics: Martin D. Kruskal, Rutgers University, for discovery of the soliton and research on nonlinear equations;

Physical sciences: Val L. Fitch, Princeton University, for pioneering physics research and national service; Vera C. Rubin, Carnegie Institution, for research in observational cosmology.

National Medal of Technology

Advanced manufacturing: Walter L. Robb, General Electric;

Human resource development: Hans Liepmann, California Institute of Technology, aeronautical engineering;

Product and process innovation: Amos E. Joel Jr., AT&T Bell Labs; William H. Joyce, Union Carbide; George Levitt, DuPont; Marinus Los, American Cyanamid; and Kenneth H. Olsen, Digital Equipment Corp.;

Technology transfer: George Kozmetsky, University of Texas, Austin; and William Manly, Martin Marietta.